Onechte variabelen in de ekonometrie

Kooyman, Marius Anthonie

Summary

Econometrics deals with variables the quantitative character of which may arise in two different ways. Firstly, use is made of quantitative variables which represent measurable economic magnitudes, such as income, price, unemployment, imports, exports etc. Secondly, quantitative variables are used to indicate attributes, e.g., sex, social status, educational level, race, etc. These two types of variables will be referred to as „real" and „dummy" variables respectively. The use in econometrics of dummy variables in the above sense is the topic of this study.

Within the class of dummy variables, the binary or zero-one variables play a dominant part in econometrics. The main part of this study is concerned with binary variables. Such a variable can be applied to indicate the presence or absence of some attribute, e.g. by allotting „*1*" to wartime years and „*0*" to peacetime years or „*1*" to males and „*0*" to females.

Chapter II analyses the meaning of regression coefficients of binary variables used as regressors in a linear relation. It is shown that if such variables are used together with real variables, the corresponding regression coefficients of the binary variables can be expressed as linear combinations of certain class averages of the real variables involved. In the special case where the regression coefficients of the real variables do not change with shifts in the binary variables, the regression coefficients of the binary variables indicate shifts in the constant of the linear relation.

Dummy variables which can assume more than two values, e.g. the values *0*, *1* and *2* are seldom applied in econometric publications. This study shows that the arbitrariness of the values chosen for such dummy variables (which in fact represent unmeasurable phenomena) has its unfavourable effects on all the regression coefficients of the linear relation, which become uncertain or even meaningless. Binary variables do not entail this drawback. It is argued on other grounds also that multi-valued dummy variables should be avoided in econometric analysis.

This Chapter further considers the situation of a binary regressand. It is shown that such a regressand does not fit very well in the linear regression

erros in explanatory variables. A generalized version of Wald's estimator of the regression coefficient is derived on the basis of the introduction of a binary variable into the estimation procedure according to the two-stage-least-squares method.

Chapter III presents a detailed account of the relations between regression analysis using binary variables as regressors on the one hand and analysis of variance and covariance on the other. Tests pertaining to variance- and covariance analysis are shown to be equivalent in this context to some well-known regression tests, such as the test of significance of R^2 and the t- and F-tests of significance of a number of regression coefficients.

In Chapter IV binary variables are introduced to enable the research worker to estimate period-to-period shifts in regression coefficients. It is shown that Chow's test of equality between sets of regression coefficients in two linear relations can be reformulated as a well-known regression problem, namely that of testing the significance of a number of regression coefficients. Tackling the problem of testing the constancy of regression coefficients over time, if carried out with the help of binary variables, opens up a wider perspective than Chow's analysis. This new method even analyses the constancy of the coefficients of simultaneous equation systems. The strict assumptions underlying Chow's method can furthermore be released considerably.

The theoretical results obtained in Chapter IV are applied in Chapter V. In this last chapter the constancy of the coefficients of a large economic model for the Netherlands (the so-called CPB-model 63-D), constructed by the Central Planning Bureau, is investigated by the introduction of a number of binary variables. The coefficients of this model, estimated by the Bureau, are based on observations over the periods 1923-1938 and 1949-1962. It is demonstrated in Chapter V that the hypothesis maintained by the Central Planning Bureau, namely of equality of all corresponding prewar and postwar coefficients, should be rejected. In fact, approximately 15% of these coefficients have undergone a statistically significant change between the prewar and the postwar period.